NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

SPEECH RECOGNITION AND THE
TELECOMMUNICATIONS EMERGENCY DECISION
SUPPORT SYSTEM

by

Nancy C. Browne
MARCH 1991

Thesis Advisor: Daniel R. Dolk
Co-Advisor: Gary K. Poock

Approved for public release: Distribution is unlimited

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0188

1a. REPORT SECURITY CLASSIFICATION: Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release: Distribution is unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION: Naval Postgraduate School — 6b. OFFICE SYMBOL (If applicable): AS	7a. NAME OF MONITORING ORGANIZATION: Naval Postgraduate School
6c. ADDRESS (City, State and ZIP Code): Monterey, CA 93943-5000	7b. ADDRESS (City, State, and ZIP Code): Monterey, CA 93943-5000
8a. NAME OF FUNDING/SPONSORING ORGANIZATION — 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBER: PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

SPEECH RECOGNITION AND THE TELECOMMUNICATIONS EMERGENCY DECISION SUPPORT SYSTEM

12. PERSONAL AUTHORS

NANCY C. BROWNE

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Master's Thesis	FROM ____ TO ____	MARCH 1991	54

16. SUPPLEMENTARY NOTATION

The views expressed are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block numbers)
FIELD	GROUP	SUB-GROUP	Speech recognition

19. ABSTRACT (Continue on reverse if necessary and identify by block numbers)

The purpose of this thesis is to provide a feasibility study for incorporating speech recognition into the Telecommunications Emergency Decision Support System (TEDSS) developed by the National Communications System (NCS) and contained on a Compaq 386. The three types of speech recognition systems that were used are: the DragonDictate, a software driven system, the Verbex Series 5000, a system contained in a peripheral device, and the KeyTronic Speech Recognition System, a system contained in a keyboard in addition to using speech software. A prototype was developed using the speech systems to determine whether or not TEDSS could be combined successfully with speech recognition. The results indicate that the incorporation of speech recognition into TEDSS is possible with some modifications to TEDSS software and to the Compaq 386.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: XX UNCLASSIFIED/UNLIMITED _ SAME AS RPT _ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION: unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: Daniel R. Dolk	22b. TELEPHONE (Include Area Code): (408) 646-2260 — 22c. OFFICE SYMBOL: AS/DK

DD Form 1473, JUN 86 Previous editions are obsolete SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603 Unclassified

Approved for public release: Distribution is unlimited

Speech Recognition and the Telecommunications Emergency Decision Support System

by

Nancy C. Browne
Captain, United States Army
B.A., Northeastern University
M.S.B., Troy State University

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN
INFORMATION SYSTEMS

from the

NAVAL POSTGRADUATE SCHOOL

MARCH 1991

David R. Whipple, Chairman
Administration Sciences

ABSTRACT

The purpose of this thesis is to provide a feasibility study for incorporating speech recognition into the Telecommunications Emergency Decision Support System (TEDSS) developed by the National Communications System (NCS) and contained on a Compaq 386. The three types of speech recognition systems that were used are: the DragonDictate, a software driven system, the Verbex Series 5000, a system contained in a peripheral device, and the KeyTronic Speech Recognition System, a system contained in a keyboard in addition to using speech software. A prototype was developed using the speech systems to determine whether or not TEDSS could be combined successfully with speech recognition. The results indicate that the incorporation of speech recognition into TEDSS is possible with some modifications to TEDSS software and to the Compaq 386.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. INTRODUCTION

A. BACKGROUND

The National Communications System (NCS) is responsible for coordinating national and regional telecommunication resources in case of a national emergency of any type. To meet this responsibility, NCS has developed a decision support system called the Telecommunications Emergency Decision Support System (TEDSS) to assist in the management of telecommunication resources on a national level. TEDSS will be used in times of national emergency by regional managers who may not have a high degree of computer expertise.

B. THE PROBLEM

TEDSS provides automated, interactive information processing and decision support to NCS in times of national emergency. The eventual users of TEDSS will be "computer naive" regional managers operating under time constraints in an emergency situation. As a result, they may be reluctant to use a keyboard to interact with TEDSS since it would require time they are not willing to relinquish. Speech recognition is a technology which can reduce the time and complexity of interaction and potentially increase TEDSS' usefulness. If speech recognition can be combined with TEDSS, the system may

be more accessible and user friendly under emergency conditions.

C. SPEECH RECOGNITION TECHNOLOGY

The role of speech recognition in desktop computing is not as well established as in manufacturing, inventory control, etc. where the user's hands and eyes are otherwise occupied. However, the success of speech recognition is predicated on our understanding of what it can and cannot do as it evolves. The critical tests of practicality, reliability, user desirability, and cost effectiveness may be met for a number of applications by today's products. Nevertheless, more understanding of the unpredictable human element must be achieved. Research is currently attempting to do this. It is only by continuing research and development with automatic speech recognition that we can define and refine the work remaining to realize its full potential.

D. METHODOLOGY

Three types of speech recognition systems were tested. Each represented a different approach to incorporating speech recognition with TEDSS. The first was the DragonDictate by Dragon Systems, Inc., a software driven speech system using a speech processor board installed in a Compaq, and a head microphone which pluged in to the speech processor board. This software was used to test and verify the speech system's

ability to operate a menu-driven application such as TEDSS. The second system was the Verbex Series 5000, by Verbex Voice Systems, which is completely self-contained in a peripheral device. The system represents a hardware alternative to the first approach and requires significantly less hard disk space. The third was the Key Tronic Speech Recognition Keyboard, by KeyTronics, which uses a keyboard as an external device along with the speech software. The speech processor is contained within the keyboard and uses a head microphone which plugs into the keyboard. This alternative was used as a compromise between having the speech system either totally contained internally or contained externally in a peripheral device. Each system was initially tested as a standalone system for familiarization and to determine ease of training. Upon completion, attempts were made to incorporate each system into TEDSS.

E. SCOPE OF THE PROBLEM

This thesis examines and evaluates each of the three types of speech recognition systems based on their interaction with TEDSS software and the Compaq hardware. Since TEDSS will be used in emergency situations, evaluation criteria that were considered in addition to operational capability include portability, ease of training, and installation requirements, if any.

F. STRUCTURE OF THE THESIS

This thesis will review TEDSS and its architecture, current speech recognition technology, and the development of a prototype combining the two. The prototype is used to determine the feasibility of whether or not TEDSS can be combined successfully with speech recognition. Problems resulting from design constraints within TEDSS are identified and addressed along with any hardware constraints within the Compaq. Recommendations for resolution of these problems are included along with suggested areas of research for future theses.

II. TEDSS ARCHITECTURE AND CAPABILITIES

A. BACKGROUND

The purpose of TEDSS is to provide automated, interactive decision support to the Office of Manager, NCS, (OMNCS) for the management of national telecommunication resources in times of national emergency, and to support the six federal regions for the management of regional resources. Since user requirements at the national and regional levels are different, the TEDSS operational configuration is divided accordingly. The national component deals with high level information regarding the management of telecommunication resources on a national level, while the regional component is primarily involved with detailed information about regional telecommunication assets.

The national data resides at the designated National Communications Center (NCC) while copies of regional data bases are kept on the regionally deployed TEDSS. Each region is required to be able to assume the duties of the NCC, consequently a backup copy of the national data base is contained on each regional system. However, the OMNCS retains control of the update, deletion, and maintenance of the national data base. A regional user can access the national

data base using any of the three following methods, each with its own login and password.

- Regular Operations: day-to-day non-emergency operations
- What-If: allows regional managers to participate in regional exercises or game-playing. Here the user is allowed to change the national data base but only on a temporary basis. The national data base is later restored to its original state.
- Emergency: under emergency conditions, the regional manager assumes the role of the national manager and has full read and write access to the national data base.

B. SYSTEM FUNCTIONS

There are two versions of TEDSS: one version running on a MicroVax II and the other, a "portable" version which runs on the Compaq 386. Both versions use the Unix operating system. Unix is a multitasking operating system that allows a user to initiate multiple tasks, run them concurrently, and switch freely among them. Access to TEDSS functions and data is controlled through the use of log on and password capabilities. Upon activation, the system automatically requests the user to log on and enter the password. There is no interaction between the user and the Unix operating system outside of TEDSS. Interaction with TEDSS is accomplished through menu-driven software that allows the user to move within a hierarchy of menus. (See Figure 1.) TEDSS provides the user with an on-line help facility to assist with run-time operation of the system. Text defining system operation and

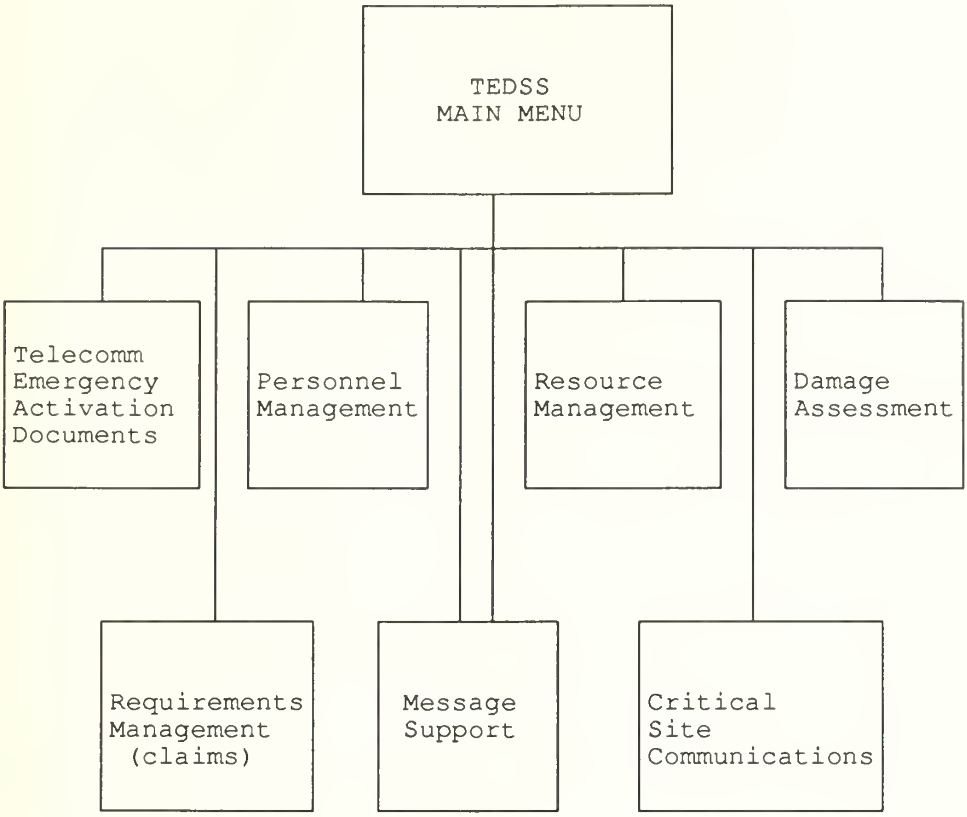

Figure 1. TEDSS Main Menu

commands is displayed with prompts to allow for continuation screens. The software supports each of the following seven major functional areas:

1. Telecommunications Emergency Activation Documents
2. Personnel Management
3. Resource Management
4. Damage Assessment
5. Requirements Management (claims)
6. Message Support
7. Critical Site Communications

Special function keys are provided to facilitate manipulation of the screens, prevent accidental corruption of data, and assist the user in moving between the various functions. The purpose of each of these keys is displayed and include: movement around the TEDSS menu hierarchy, a help facility, a print screen, and data update authorization.

1. Telecommunications Emergency Activation Documents

This function has the capability to retrieve and display the Office of Science and Technology Policy (OSTP) Telecommunication Orders (TELORDS), the NCS Telecommunication Instructions (TELINSTR), and the Presidential Executive Action Documents (PEAD). (See Figure 2.)

These documents contain predefined instructions on the roles and responsibilities of the OMNCS during a state of national emergency. This function also allows the user to review and update both the overall current status of the nation's state of emergency and the current status in each of the following six Federal Regional Center: Maynard, Massachetts; Thomasville, Georgia; Denton, Texas; Battle Creek, Michigan; Denver, Colorado; Bothell, Washington.

2. Personnel Management

This option provides a list of all personnel to be contacted in the event of an emergency such as, points of contact for the emergency operation center and for various

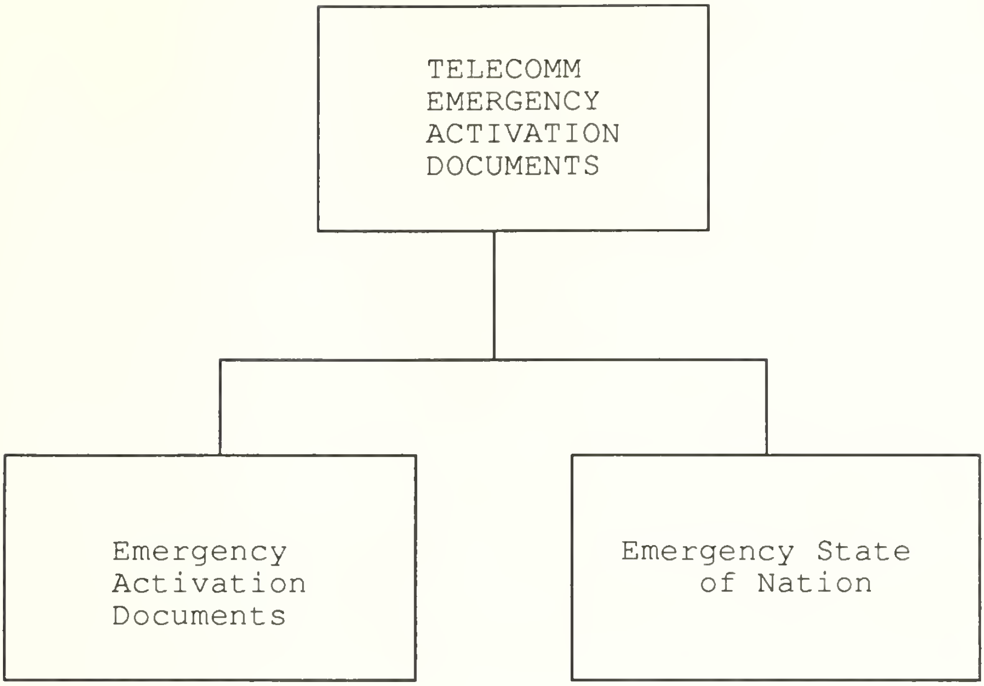

Figure 2. Telecommunications Emergency Activation Documents

telephone companies. The user can update or delete the information as necessary.

3. Resource Management

This function enables the user to update and monitor national telecommunication resources. (See Figure 3.) These resources are categorized as: Personnel, Networks, Nodes, Links, Operations Center, Asset Centers, and Assets (general). Based on parameters selected by the user, telecommunication resources within an area are displayed in a standard format. The locations of the resources can be displayed on a map of the nation by federal region or by state. The parameters can

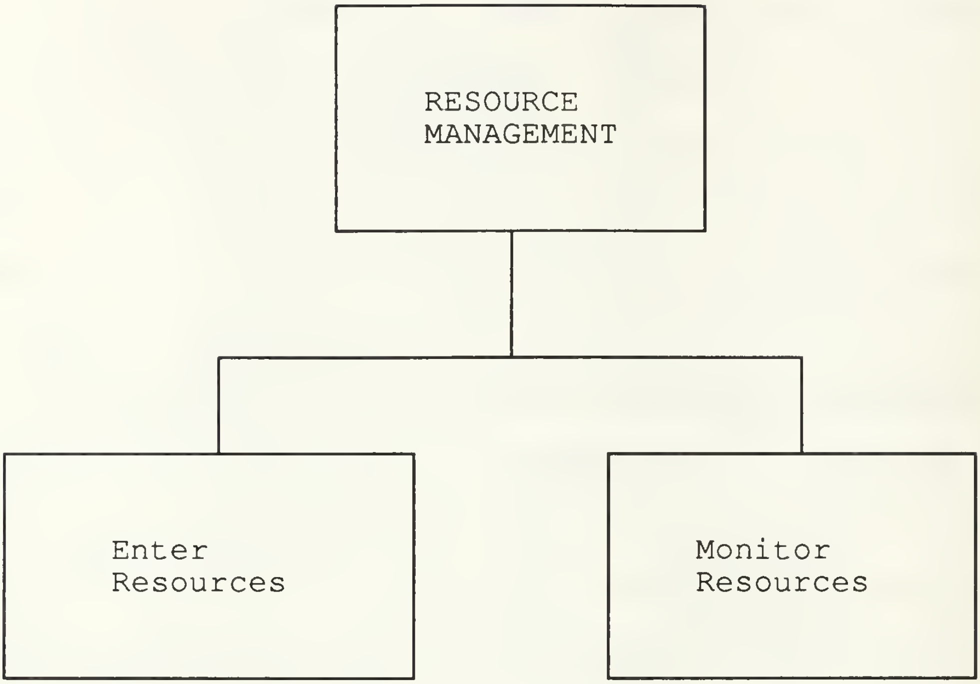

Figure 3: Resource Management

be changed in order to adjust the display. If desired, all information on a specific resource can be retrieved and displayed and, if necessary, updated.

4. Damage Assessment

This is a damage assessment model which simulates a nuclear attack. It enables the user to identify telecommunication resources that may have been damaged in a nuclear attack. (See Figure 4.)

When the location and extent of the damage are provided to TEDSS, the status of telecommunications resources affected will be updated to either predicted impaired or predicted destroyed. Each report will contain a summary of the

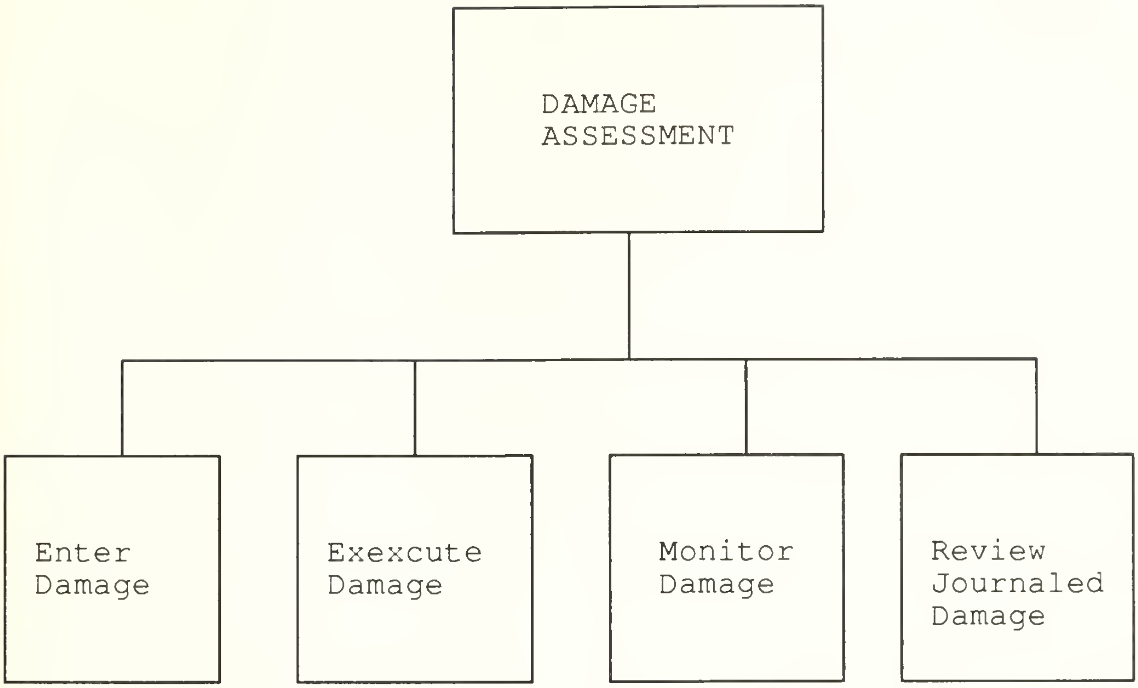

Figure 4. Damage Assessment

impact of an emergency on the telecommunications resources in the affected area. The assessment capability allows the user to update, execute all of the damage information in the TEDSS data base against all resources, monitor damage to locations and telecommunications resources, and review damage that has been entered into an on-line journal. Damage reports can be provided summarizing the impact on the resources by region or by state and type. If needed, a graphical representation of the damaged resources in a particular area can also be provided. Any damage information which is no longer valid may be sent to a Damage Journal where it may be edited and mapped, or deleted.

5. Requirements Management (Claims)

Allows the user to enter a request for restoration or augmentation of existing failed telecommunications services such as telephones, networks, switches, microwave, etc. (See Figure 5.)

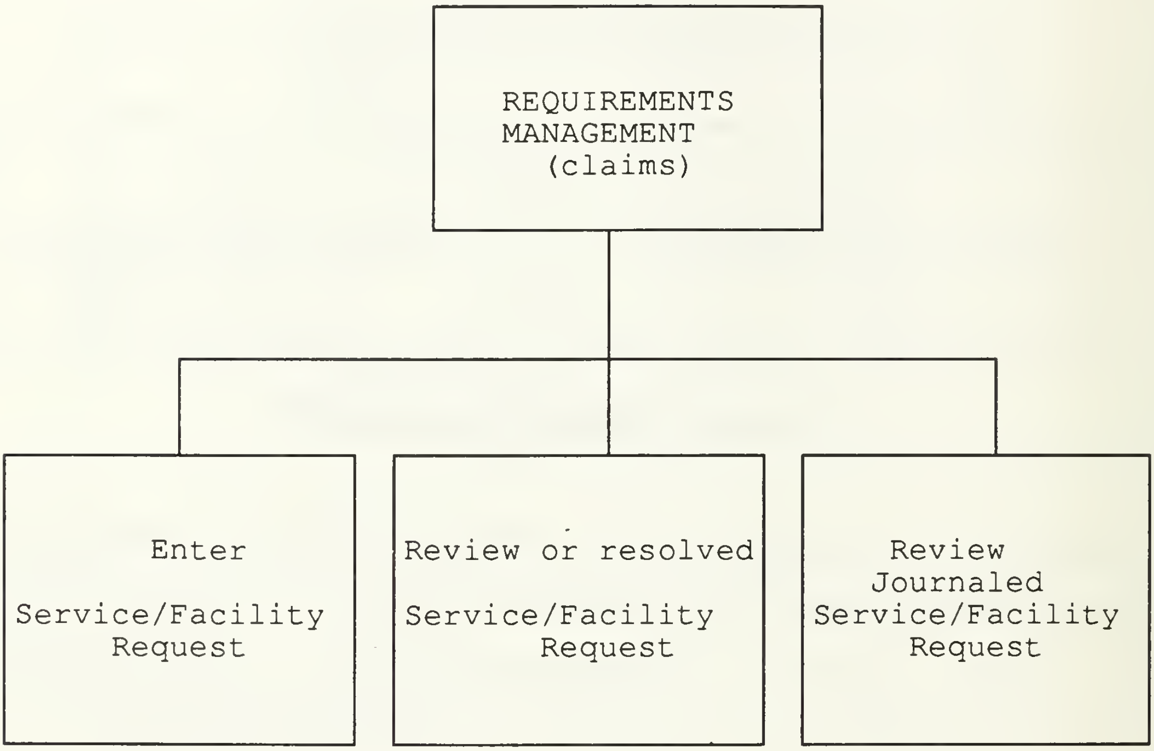

Figure 5. Requirements Management

a. *Enter a service or facility request*

All requests from NCS member agencies may be entered into the data base utilizing a standard format provided by the system. TEDSS assigns a unique NCC number to each request, and all requests are maintained in a prioritized order based on predetermined factors.

b. Review and resolve service or facility requests

This function enables the user to review, edit, and update requests, or resolve claims for service or facilities on any active requests by providing a point of contact for resolving a claim. Once resolved, the claim and its resolution are entered into the system's journal.

c. Review journaled service or facility requests

This option reviews service or facility requests that have been moved from the active list of requests. These requests can still be edited or deleted, as appropriate.

6. Message support

TEDSS provides interactive communication between two users enabling them to send and receive information simultaneously through the phone option. (See Figure 6.)

Non-interactive communication allowing users to send mail to other users of the system is provided through the mail option. Upon logging in to the system, a user is notified of any mail received.

7. Critical site communication

This function provides the national manager, or the regional manager acting as the national manager, the ad hoc ability to input engineered networks, and generate a new network. (See Figure 7.)

It enables the manager to identify and establish communication between two critical persons or locations. It

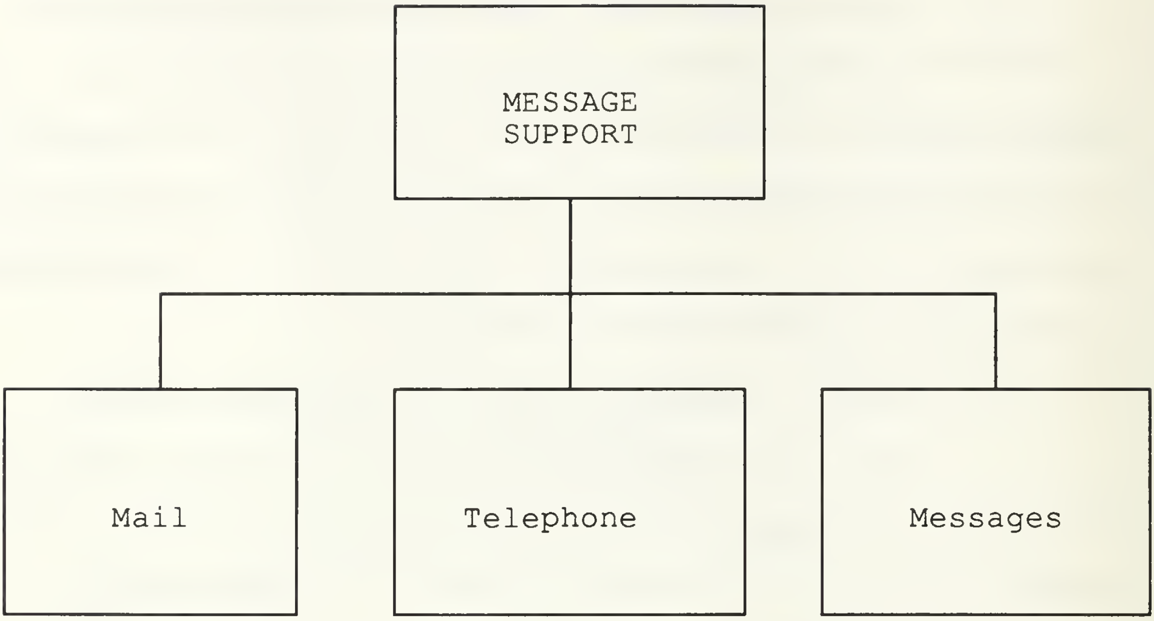

Figure 6. Message Support

also lists all on-line systems where communication has been established.

C. HARDWARE

The national level component of TEDDSs is on a MicroVAX II minicomputer which contains the data base in disk storage manipulated by the INGRES data base management system. The MicroVAX II, a Digital Equipment Corporation (DEC) computer system, uses the VAX/VMS operating system which is a general purpose operating system. It provides a reliable, high performance environment for the concurrent execution of multi-user timesharing, batch and real-time applications. There are several terminals directly connected to the MicroVAX along

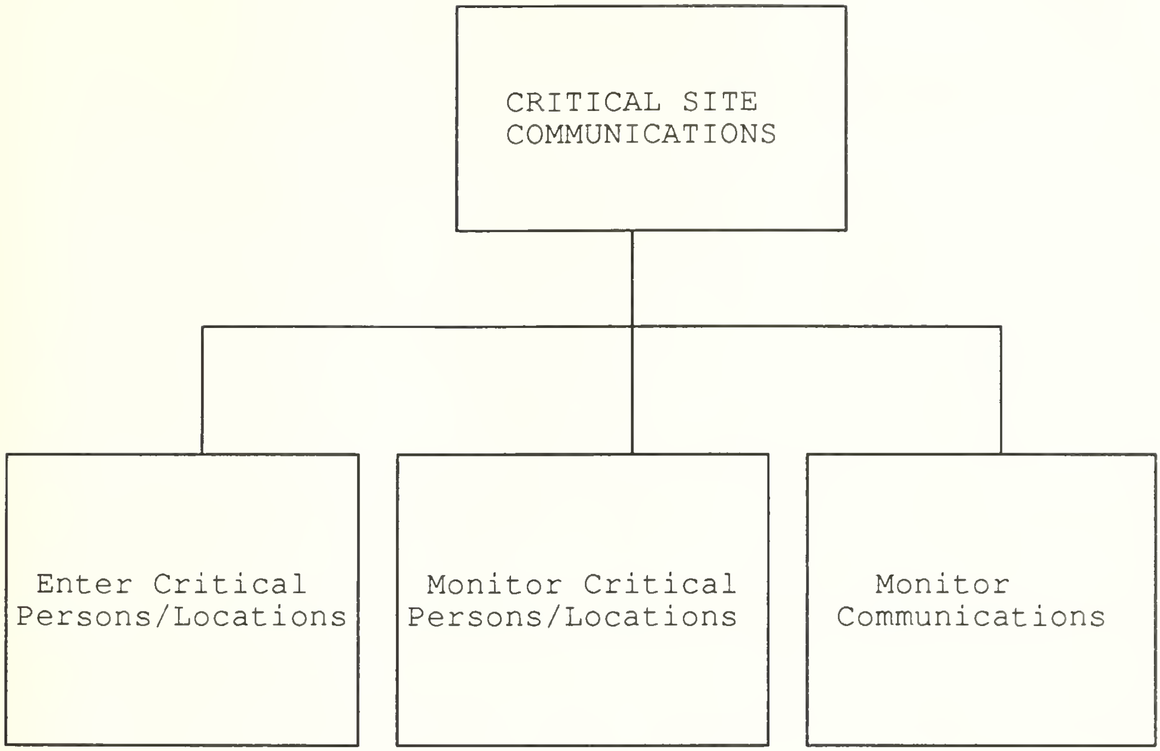

Figure 7. Critical Site Communications

with a magnetic tape drive for back-up and archiving, and a line printer for hard copy reporting. (See Figure 8.) The communications interfaces for the peripheral devices and external communications interfaces are also on the MicroVAX II.

The regional TEDSS operating environment is essentially the same as that on the national level. The personal computer used is a Compaq portable 386 linked to a DEC MicroVAX. The TEDSS software is on the MicroVax while the graphics module and the PC/VAX communications software is on the Compaq. The

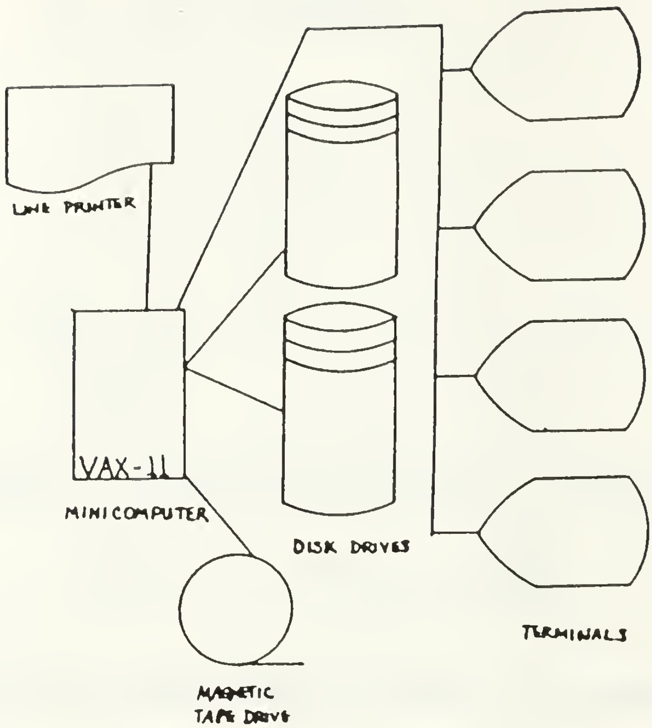

Figure 8. MicroVAX II Configuration

regional components communicate with each other and with the national node via the DECNET communications network.2

III. CURRENT SPEECH RECOGNITION TECHNOLOGY

A. BACKGROUND

For a long time, interaction between voice and computing, which can take many forms, has been categorized under the general heading of voice/data integration. This narrow designation usually implies the existence of several digital information streams, some representing voice content and some containing data, which have been multiplexed into a single physical channel. In reality, the range of available technology supporting the interaction of voice and computing is more diverse. Voice technologies can be separated into three general categories: connection control, and software architecture and content processing. Connection control is the arrangement of voice channels to interconnect users and voice equipment. It includes telephone signaling arrangements and point-to-point command links. Software architecture is the organization of computing system software to facilitate the creation of voice-related applications. It includes the abstract modeling of voice resources and distributed access to voice resources. Content processing is the creation, manipulation, and analysis of the information appearing in a voice channel. Speech recognition is included in this category

and, for our purposes, we will limit this discussion to speech technologies only.

Speech recognition is the capability of recognizing spoken utterances from a given vocabulary set. There are approximately 43 distinct sounds that make up our spoken language. These sounds, known as phonemes, comprise a set of distinct, mutually exclusive speech sounds that may be found in almost any spoken language. These phonemes are distinguishable from each other primarily by the range of frequencies generated by the vocal tract during their production. The air passages above the vocal cords are known collectively as the vocal tract. It extends from the larynx or "voice box" to the lips and includes the entire area of the mouth. The vocal tract acts as a resonant "hole" or hollow area intensifying certain frequencies and weakening others. As speech is generated, the initial sound comes from a vibration in our vocal cords. This sound is generated by the vocal cords rapidly opening and closing with small puffs of air.

Some of the phonemes belong to a group called continuants which are sustained sounds such as vowels. These phonemes, because of a lack of vocal tract motion during speech, have a stable and constant frequency range throughout their vocalization. Other classes of phonemes are the plosives and the glides. Plosives are produced by the complete stopping and sudden release of the breath such as "b" in base. The glides are sounds that flow, such as "y" in you. Both plosives and

glides are considered to be sounds that normally couple to the surrounding phonemes in a manner resembling diphthongs. Diphthongs exist as a class of speech sounds characterized by extreme vocal tract motion when coupling other phonemes together. They are generated as the mouth moves from one phoneme position to the next during speech, such as the "g" in get or the "w" in will. Since the response time of the muscles within our throat and mouth tend to slur the movement from one spoken phoneme to the next, many diphthongs are generated within our speech patterns.

Although the number of phonemes is small, their automated recognition by a computer system is still a problem since only recently have there been well-defined sound patterns or templates for phonemes. Each phoneme has a different duration, and certain vowel sounds can be assigned equally to different phonemes. However, improved technology in phonetic recognition has recently achieved greater degrees of success and higher recognition rates. The phoneme patterns of a language are limited not only by the set of sounds themselves, but also by the allowable combinations. By incorporating rules based on the allowable phoneme combinations in a phonetic recognizer, more robust speech recognition front-ends can be built. The emphasis in speech recognition has been on pattern-matching of word-sized units with those already stored in the data base. The problems associated with finding the best match, and insufficient speed of digital processing, have hindered

progress in this area. Parallel processors and intelligent algorithms that use parallel architectures fully should help to resolve these problems.

B. TYPES OF SPEECH

The most general forms of speech recognition are speaker-dependent, speaker-independent, discrete speech and continuous speech.

A speaker-dependent system requires that samples of the user's voice be in memory in order to work properly. Since this system is basically tuned to a particular user's voice, it is easier to recognize than speech which may originate from a variety of speakers. The parametric representations of speech are sensitive to the characteristics of a specific speaker. This makes a set of pattern-matching templates for one speaker perform poorly for another speaker. Consequently, many systems are speaker-dependent, trained for use with each different user.

A speaker-independent system contains algorithms which can handle many different voices and diale s. Because of these robust algorithms, the system should be able to recognize the voice of anyone who tries to use it.

In a discrete speech system, the user has a given number of sound patterns in memory. A sound pattern can be one or several words in a continuous phrase of sound. When using the discrete system, a user must pause about .10 seconds between

each utterance made. When the system 'hears' the pause, it knows that was the end of an utterance and therefore starts to search the memory for what was just said. In a continuous speech system, no pause between utterances is required. It is the job of the recognition algorithm to determine word boundaries. Also, coarticulation effects in continuous speech can cause the pronunciation of a word to change depending on its position relative to other words in a sentence. Coarticulation is a dependence on the preceding sounds and anticipation of the following sounds. For example the statement, "What did you do last night?" can become, "Whajedolasnigh?"

Additional factors affecting speech recognition are vocabulary size, grammar, and environment. The size of the vocabulary of words to be recognized also influences recognition accuracy. Large vocabularies are more likely to contain ambiguous words than small vocabularies. Ambiguous words are those whose pattern-matching templates appear similar to the classification algorithm used by the recognizer, consequently they are harder to distinguish from each other.

In the recognition domain, grammar defines the allowable sequences of words. A tightly constrained grammar is one in which the number of words that can legally follow any given word is small. The amount of constraint on word choice is known as the perplexity of the grammar. Systems with low

perplexity are potentially more accurate than those that give the user more freedom. The system can limit the effective vocabulary and search space to those words that can occur in the current input context. Background noise, changes in microphone characteristics, and loudness can all dramatically affect recognition accuracy. Many recognition systems are capable of very low error rates as long as the environmental conditions remain quiet and controlled. However, performance degrades when noise is introduced or when conditions differ from the training session used to build the reference templates. To compensate, the user must almost always wear a head-mounted noise-limiting microphone with the same response characteristics as the microphone used during training.

C. CURRENT SYSTEMS

Current speech recognition systems can be divided into two primary categories: speaker-independent or speaker-dependent. A summary of the capabilities, costs, and manufacturers' claimed accuracy of a sample of commercial products of current systems representing these categories are presented in Table I.

The DragonDictate shown in Table I represents a category in speech recognition systems known as speaker-adaptive. The user's speech is not required to be in memory prior to operating; however, it "learns" and adapts to the voice of the user with each successive use. The system recognizes 30,000

TABLE I. EXAMPLES OF SPEECH RECOGNITION SYSTEMS

System	Constraints	Price	% Word Accuracy*
ITT VRS 1280/PC	Spkr-Depnd Continuous Speech 2,000 words	$9,000	>98
Phonetic Engine (Speech Systems, (Inc)	Spkr-Indep Continuous Speech 10,000-40,000 words	$10,500-$47,100	95
Verbex Series 5000, 6000, 7000	Spkr-Depnd Continuous Speech 80-10,000 words	$5,600-$9,600	>99.5
Voice Card (Votan)	Spkr-Depnd/Indep Continuous Speech 300 words	$3,500	>99 (Depnd) 95 (Indep)
Voice Navigator (Articulate Systems)	Spkr-Depnd Isolated-word 1,000 words	$1,300	95
Voice Report (Kurzweil AI)	Spkr-Depnd Isolated-word 20,000 words	$18,900	98
DragonDictate (Dragon Systems)	Spkr-Adaptive Isolated-word 30,000 words	$9,000	>90

*As claimed by vendor

words or utterances surrounded by brief pauses of .25 seconds. This is slower than discrete speech which usually has pauses of .10 seconds. The 30,000 words is a soft limit. After reaching this limit any time a new word is used, the word least recently used will be deleted from the vocabulary. In

this way, the system constantly adapts to the changing vocabulary.

D. USES IN INDUSTRY

Speech recognition through the telephone system is particularly useful, since hundreds of millions of telephones are in use today. Equipped with speaker-independent speech recognition and synthesis equipment, a computing application can use these telephones as input/output devices, making all telephone subscribers potential users. Voice interaction will allow people to communicate directly with computers to perform simple tasks without the need for operators. Automating the telephone operator's job by using interactive voice technologies can greatly reduce operating costs for telephone companies and provide a host of new services for consumers. It may put some people out of work, however.

Speech recognition is currently being applied most often in manufacturing for companies needing voice entry of data or commands while the operator's hands are otherwise occupied. Related applications are product inspection, inventory control, command/control, and material handling. In the medical field voice input can significantly increase the writing of routine reports. In Japan, Nippon Telegraph and Telephone has combined speaker- independent speech recognition and speech synthesis technologies in a telephone information system called ANSER (Automatic Answer Network System for

Electrical Requests). ANSER's voice response and voice recognition capabilities let customers make inquiries and obtain information through a dialogue with a computer. However, speaker-independent speech recognition is particularly difficult through telephone lines because, in addition to the variations among speakers, telephone sets and lines cause varying amounts of distortion. To simplify the manipulation of speech data, ANSER has incorporated several original modifications of conventional speech recognition and synthesis technologies.

Being able to speak to your personal computer, and have it recognize and understand what you say would provide a comfortable and natural form of communication. It would reduce the amount of typing required, and leaves the hands free for other tasks. Forms of speech recognition are available on personal workstations. With the current interest in speech recognition, performance of these systems is improving. Speech recognition has already proven useful for certain applications, such as telephone voice-response systems for selecting services or information, digit recognition for cellular phones, and data entry while walking around.

The role of speech recognition in desktop computing is not so well established as in manufacturing, inventory control, etc. where the user's hands and eyes are otherwise occupied. Researchers at the Massachusetts Institute of Technology have focused on window systems, where speech might provide an

additional channel to support window navigation [Ref. 1]. Xspeak, their speech interface to the X Window System, associates words with each window. By speaking a window's name, it is moved to the front of the screen and the cursor is moved into it. Speech does not provide a keyboard substitute, but it does assume some of the functions currently assigned to the mouse. Consequently, a user can manage a number of windows without removing his or her hands from the keyboard.

Past work at Boeing in voice-controlled computer applications included a robotic vocational workstation for the physically disabled professional [Ref. 2]. Through voice commands and a specially designed robotic arm, users could retrieve documents from a printer, pick up books, and perform other manipulative tasks. A voice-operable telephone management system allowed users to receive telephone calls, record notes and incoming messages, create phone number indexes and directories, and access on-line databases and bulletin boards. The workstation could be connected to various network systems allowing users to access information from remote computer sites by voice. Users activated and shut down their workstations by moving their wheelchairs to break a light beam underneath their desks.

IV. DEVELOPMENT OF THE PROTOTYPE

A. HARDWARE

The portable version of TEDSS is contained on a Compaq 386 computer with 110 megabytes of hard disk and ten megabytes of RAM. It is a menu-driven application that operates under the UNIX operating system utilizing UNIX configuration and commands. A Unix feature, the VP/IX, provides an emulation of MS-DOS. Its main purpose is to allow applications that were developed under MS-DOS to run as Unix processes. The organization of tree-structured directories is identical in MS-DOS and in Unix. Consequently, one can move between directories using similar commands. Since it is possible to run MS-DOS as a session under Unix 286, 386, and 486 machines, the consistency of file structure allows manipulation of files from both operating systems. Although Unix is the primary operating system on the Compaq, it contains an MS-DOS partition. A partition is a self-contained area of the hard disk with boundaries that separate it from other partitions. Within the MS-DOS partition are application programs, such as WordPerfect and MapInfo, that require the MS-DOS operating system. (See Figure 9.)

The hard disk on the Compaq is separated into two partitions. The first partition contains 100 megabytes with

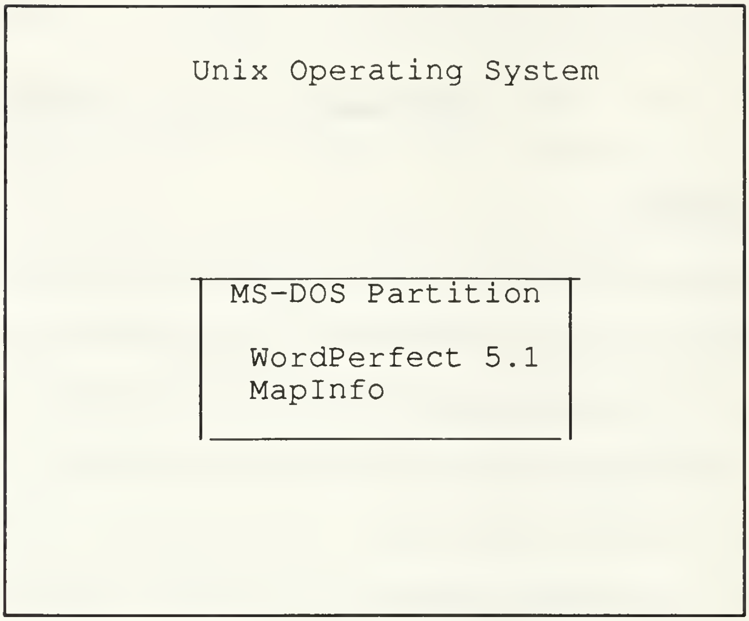

Figure 9. MS-DOS Partition

Unix using approximately 80%. The second partition contains 10 megabytes with the MS-DOS partition using approximately 8.5%. The Compaq also contains 10 megabytes of RAM. TEDSS is designed so that upon start-up, it automatically puts the user into the application. Consequently, because of this tight design, and its utilization of 80% of its partition, there is no room for additional applications to be loaded within the Unix configuration.

B. THE SPEECH RECOGNITION SYSTEM

Speech recognition systems are operated by either loading the speech software into the system and installing a speech board containing a speech processor, or by plugging into the

serial port a peripheral device which contains the speech processor. One system that could be used for TEDSS is the DragonDictate by Dragon Systems, Inc, a state-of-the-art speaker-dependent, discrete system which can recognize up to 30,000 words at a time and has access to an 80,000 word on-line Random House Dictionary.

The DragonDictate system is composed of three high density 5 1/4" floppy disks containing the speech recognition software and the word library, a speech board containing the speech processor, and a head-mounted microphone which plugs into the speech processor board. The speech processor has been designed to use voice commands, keystrokes, or any combination of voice and keystrokes. Any functions that can be handled by the keyboard can now be handled by voice commands. It requires MS-DOS version 3.3 or higher, an 80386 based computer that is PC/AT or PS/2 compatible system, either 6 megabytes of RAM for start-up or 8 megabytes of RAM for full vocabulary access, a hard disk with a minimum of 8 megabytes of free disk space, and a high density floppy drive. Each additional user who creates a file of their voice patterns will require an additional 2.5 megabytes. Currently most of the manufacturers of speech recognition systems operate using the MS-DOS operating system and have no immediate plans for interfacing with UNIX. However, ITT Corporation does have a speech system which runs on the Xenix operating system and is compatible

with Unix, but Xenix is not used in TEDSS. Also, the ITT system is quite expensive with a purchase price of $12,000.

C. METHODOLOGY

1. The DragonDictate

Based on its operating system requirements, the DragonDictate was loaded into the MS-DOS partition. It is fully operational in the partition and, once samples of the user's speech pattern are in memory, is able to recognize the user's speech. With DragonDictate the user can activate and operate any application within the partition such as WordPerfect 5.1. The multitasking feature of Unix is activated through the MS-DOS emulator, the VP/IX. It contains the batch files for the applications within the MS-DOS partition. Batch files are files that contain the sequence of instructions and the command of execution for a specified application. Once DragonDictate has been activated within the partition by the batch file, the user must be able to access the TEDSS main menu from the Unix operating system. However, TEDSS is not designed for interaction between the user and the operating system. Consequently, without a bridge or command channel between Unix and TEDSS, the multitasking feature which would enable TEDSS to access the DragonDictate under the VP/IX shell is inoperable. DragonDictate itself works fine and there would be no problems using the Dragon system on the TEDSS if, and when the multi-tasking feature ever becomes operable. Research

should continue in developing the vocabulary to be used with TEDSS in the future.

2. KeyTronic Speech Recognition Keyboard

Since TEDSS is designed to accept input from the keyboard, an alternative approach considered was the KeyTronic Speech Recognition Keyboard. The KeyTronic speech recognition speech processor is contained within the keyboard. The layout of the keyboard is basically unchanged since the head-mounted microphone plugs directly into the rear of the keyboard. However, since the Compaq comes with the keyboard attached, a simple adaptor needs to be built to enable this type of speech recognition device to be used. The speech processor is part of the keyboard, however it's executable files are contained on floppy disks using the MS-DOS operating system. Consequently, the software which is loaded into the MS-DOS partition cannot be used to run TEDSS due to the absence of a command channel between Unix and TEDSS. TEDSS could run with KeyTronic speech input, however an access input must be provided for the speech signal to the TEDSS system. In the meantime, research should continue to develop the actual vocabulary now needed to operate TEDSS.

3. Verbex Series 5000

Another approach was the Verbex Series 5000, a speech recognition system completely self-contained in a peripheral device. The Verbex Series 5000 software and speech processor

board are contained within a voice I/O unit which plugs into the serial port of the computer. The only external component is the head-mounted microphone which plugs into the voice I/O unit. Since there was no software to be loaded into the computer, the problem with the command channel was not applicable. However, as stated above, TEDSS is designed to accept input from the keyboard. Since the Compaq has communication capability, TEDSS has been programmed to look to the serial port for data. Therefore, the Verbex Series 5000 could not be used the way the TEDSS is presently designed, however the speech recognizer can be used to enter commands in the form of speech input. Again, the development of the vocabulary should proceed by experts familiar with speech recognition and who know how to employ speech best.

D. INTERFACE INSTRUCTIONS

If the software architecture of TEDSS is modified to make use of a speech recognition system such as the DragonDictate feasible, then the following instructions will be helpful to the System Administrator in activating the speech recognition system. When the system is turned on, a series of system checks is automatically performed. Upon completion, a Welcome screen appears requesting the system administrator to enter the proper login and password. Access to the Unix operating system is then granted and is indicated by the "#" prompt. The command "vpix" will then put the user into the DOS emulation

mode indicated by the "VP/ix Z:\>" prompt. In this mode, regular DOS commands may be used. The batch files for the DOS partition are located three levels down in the subdirectory BIN, under the subdirectory EPMIS, under the USR directory.

The following instructions describe the procedures for a user to access the DragonDictate in the DOS partition:

VP/ix Z:\> cd usr\epmis\bin [enter]

VP/ix Z:\> dir [enter]

Machine response: Lists all files in the BIN subdirectory

VP/ix Z:\> DRAGON [enter]

Machine response: Accesses the DOS partition within the Dragon directory

VP/ix D:Dragon> dt user's name [enter]

Machine response: Activates the speech recognition system

VP/ix D:Dragon> Press **[Alt-SysReq] or [Alt-SysReq-m]** (depending on the keyboard)

Machine response: VP/IX Interface Menu is displayed

VP/ix D:Dragon> R [enter]

Machine response: Reboots only the VP/IX

VP/ix Z:\> Press **[Alt-SysReq] or [Alt-SysReq-m]** (depending on the keyboard)

Machine response: Exits the emulator

#

(At this point the command to change into the established TEDSS directory can be given verbally.)

 # no space charlie delta space no space tango echo delta sierra sierra enter

Alternately, for known commands that will be needed and known ahead of time, this command could be stored as a speech phrase and one would simply say "change directory to TEDSS."

 # cd tedss

Machine response: Enters the TEDSS directory

1. Operating Within TEDSS

Following is an example of how a user could navigate through the TEDSS menu hierarchy using verbal commands. The status of where the user is within the menu hierarchy is displayed in the upper right-hand corner of each screen. The main menu displaying eight options might require the user to state the following:

```
TEST MAIN MENU

  1.   Telecommunication EADs
  2.   Personnel Management
  3.   Resource Management
  4.   Damage Assessment
  5.   Requirements Management
  6.   Message Support
  7.   Critical Site Communication
  8.   Quit

        Enter Selection:
```

"Select three" or "Resource Management" or the speech vocabulary could be working at this point where saying three would actually output a "3", or a "3 and a carriage return" as

needed. Work needs to begin on developing the vocabulary for TEDSS.

This selects Resource Management, the third option. The next level of choices within the Resource Management area is then shown.

```
                                          Main/Resources

  Telecommunication Resource Management

          1.  Enter Resources
          2.  Monitor Resources

            Enter Selection:
```

A possible voice selection to choose the second option would be:

"Select two" or "Monitor Resources" or "Two"

This command chooses the Monitor Resources option for activation. A third level of menus will appear giving the user six additional choices.

```
                                   Main/Resources/Monitor

          Monitor Resources

           1.  Networks
           2.  Nodes
           3.  Links
           4.  Operation Centers
           5.  Asset Centers
           6.  Assets

           Enter Selection:
```

A possible voice selection to choose the first option would be:

"Select one" or "Networks" or "One"

This command selects Networks as the resource to be monitored. The screen will display the following format which can then be filled in verbally by the user.

Scope:	______________
Network:	______________
Agency:	______________

Select all records that match this criteria (Y/N):

Once the form is filled in, the "Y" or "N" answer to the criterion question will automatically initiate a search of the data base based on the criteria. At any time the user may say "Select F10" to return to the previous menu shown, "Select F9" to return to the main menu, or "Select F1" to activate the help feature.

2. Summary

In order for TEDSS to work with speech input, some of the following alternatives must be implemented:

1. TEDSS must run as a separate Unix process initiated from an operating system prompt rather than running directly from login.
2. A command channel between TEDSS and Unix must be established to allow for the operation of the

multitasking feature which gives access to MS-DOS speech systems like DragonDictate under the VP/IX shell.

3. Since the Compaq comes with the keyboard attached, an adaptor can be created for the use of the KeyTronic type speech recognition keyboard.

4. Additional programming should be added to TEDSS to enable it to accept command input from the serial port.

In summary, there is no question that the TEDSS system can be run using speech input. Development of a speech vocabulary should be done immediately to prepare the TEDSS system to be used with speech input. This work can be successfully accomplished right now by building a simple adaptor to allow current ASCII signals from any speech recognizer to be passed to TEDSS on the same wiring input as the keyboard now uses. For example, splice the KeyTronic keyboard cable into the Compaq keyboard cable so that TEDSS is not aware that its commands are coming from the speech system or the keyboard. Multi-tasking, TEDSS and Unix speech systems will all be available each year in better, more advanced versions. In the meantime, development of the TEDSS vocabulary can proceed in parallel for the eventual integration of speech input with TEDSS.

V. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

It is possible to incorporate speech recognition into TEDSS at this time, but given TEDSS present design and space constraints, the operational feasibility may be a year or so away. TEDSS is a tightly designed application that requires the Unix operating system which uses approximately 80% of the 100 megabytes available in the first of two partitions. However, the use of MS-DOS as the operating system would increase the available space for additional applications. Currently, few manufacturers of speech recognition systems have future plans for developing a system that will use the Unix operating system on a personal computer. However, as Unix on PC's becomes more common, such Unix based speech systems will become available. Any non-Unix speech recognition system now used however must be loaded into the second partition using the MS-DOS operating system. Presently, 8.5 megabytes of the available 10 megabytes in the second partition are being used when applying the DragonDictate system and WordPerfect Version 5.1 thereby limiting the size of any additional software. The space requirements of DragonDictate required the removal of the MapInfo application.

TEDSS has been designed to preclude any interaction between the user and the operating system. Once the user is in TEDSS, the Unix operating system cannot be accessed by the user. Also the user, once in the operating system, cannot issue commands to change directories going from the operating system into the TEDSS directory. The reason for this is that the required programming has not been included in TEDSS software which will allow a user to change between these directories. Consequently, the programming must be modified to include a command channel between TEDSS and Unix which will contain the necessary commands. For ease of use, the programming should be structured so that the system will access the main menu upon entering the TEDSS directory. Without the command channel, once the VP/IX or Dos emulator and its multitasking feature has been activated, any speech recognition systems within the MS-DOS partition cannot be used to run TEDSS. The speech systems require access to TEDSS from the MS-DOS partition, via the DOS emulator, in order to manipulate TEDSS menu-driven software. Due to the absence of a command channel, the user currently has to reboot the system in order to enter TEDSS, thus breaking any connection established with applications in the DOS partition. TEDSS software is also written to recognize and accept input from the attached keyboard. Therefore, the hardware can be reconfigured with an adaptor to allow a speech recognition system, such as the KeyTronics keyboard which replaces the

attached keyboard, to work. For the purposes of using the internal modem, TEDSS will accept commands only from the keyboard. Consequently, additional programming must be added to TEDSS to instruct it to accept commands from other than the keyboard. This will facilitate speech recognition systems that plug in to the serial port.

B. RECOMMENDATIONS

The following recommendations are submitted:

1. It is recommended that TEDSS design be modified to allow TEDSS to run in the multitasking mode rather than as the only process.

2. Consideration should be given to either reducing the space within the first partition containing the Unix operating system in order to expand the MS-DOS partition or using MS-DOS as the primary operating system.

3. Additional programming should be added to TEDSS in order to allow it to accept input, in the form of commands, from the serial port for use of devices such as the Verbex Series 5000.

4. Reconfiguration of the keyboard attachment for the Compaq is necessary for any of the speech recognition systems that will replace the attached keyboard.

5. Proceed as soon as possible to develop the entire vocabulary of speech inputs that can be used to run TEDSS. It is only a matter of time until the details of hooking speech systems into TEDSS are solved. At that point, the vocabulary will have been developed and will be ready to go without further delay.

C. SUGGESTED FUTURE RESEARCH

Additional areas of research for TEDSS are:

1. Development and testing of a vocabulary for the TEDSS speech recognition system can be done in a lab environment at the Naval Postgraduate School (NPS). Resident expertise is available in the person of Professor Poock, an expert in speech recognition at NPS.

2. Once the vocabulary and its alternatives are developed and tested, demonstration of TEDSS and the speech input system should be done during an exercise to determine its full capability and allow for refinements. An interview of TEDSS users should be conducted to determine other ways they would like to say words/phrases to access TEDSS. Previous work by Professor Poock at NPS found, for example, eight different ways users wanted to command a system to enter a carriage return. Some alternatives were go, do it, enter, return, carriage return, get going and so on.

3. Real-time interaction between TEDSS and the Emergency Preparedness Interactive Simulation Of a Decision Environmnent (EPISODE) should be developed for use in an operational and training environment.

LIST OF REFERENCES

1. Schmandt, Chris, Ackerman, Mark S., and Hindus, Debby, Massachusetts Institute of Technology, "Augmenting a Window System with Speech Input," *Computer*, August 1990.

2. Fu, C, "An Independent Workstation for a Quadriplegic," *International Exchange of Experts and Information in Rehabilitation*, New York, 1986.

BIBLIOGRAPHY

Booz-Allen & Hamilton Inc., *Emergency Preparedness Management Information System (EPMIS): Five Year Plan (Draft)*, September 1988.

Booz-Allen & Hamilton Inc., *Emergency Preparedness Management Information System (EPMIS) Regional Component Software Requirements Specifications (Draft)*, April 1989.

Booz-Allen & Hamilton Inc., *Emergency Preparedness Management Information System (EPMIS) Software Design Specifications (Draft)*, July 1985.

Cater, John P, *Electronically Hearing: Computer Speech Recognition*, Howard W. Sams & Co. Inc., Indianapolis, Indiana, 1984.

Dragon Systems, Inc., *DragonDictate User Manual*, Newton, Massachusetts, 1990.

Fu, C, "An Independent Workstation for a Quadriplegic," *International Exchange of Experts and Information in Rehabilitation*, New York, 1986.

Lennig, Matthew, Bell-Northern Research and INRS-Telecommunications, "Putting Speech Recognition to Work in the Telephone Network," *Computer*, August 1990.

Nakatsu, Ryohei, Nippon Telegraph and Telephone, "Anser An Application of Speech Technology to the Japanese Banking Industry," *Computer*, August 1990.

Peacock, Richard D. and Graf, Daryl H., "An Introduction to Speech and Speaker Recognition," *Computer*, August 1990.

Poock, G. K, *A Longitudinal Study of Five Year Old Speech Reference Patterns*, Journal of the American Voice I/O Society, Vol. 3, June 1986.

Poock, G. K, *Experiments With Voice Input For Command and Control: Using Voice Input To Operate A Distributed Computer Network*, Naval Postgraduate School Report #NPS-55-80-016, Monterey, California, April 1980.

Poock, G. K. and Roland, E. F., *Voice Recognition Accuracy: What Is Acceptable?* Naval Postgraduate School Report #NPS-55-82-030, November 1982.

Rolands & Associates Corporation, *Users Guide for the Emergency Preparedness Interactive Simulation Of a Decision Environment (EPISODE) (Draft)*, March 1991.

Schmandt, Chris, Ackerman, Mark S., and Hindus, Debby, Massachusetts Institute of Technology, "Augmenting a Window System with Speech Input," *Computer*, August 1990.

Strathmeyer, Carl R., "Voice On Computing: An Overview of Available Technologies," *Computer*, August 1990.

Taylor, Allen, *Unix Guide For DOS Users*, Management Information Source, Inc., Portland, Oregon, 1990.

Yannakoudakis, E.J. and Hutton, P.J., *Speech Synthesis and Recognition Systems*, Ellis Horwood Limited, Chichester, England, 1987.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22304-6145	2
2.	Library, Code 52 Naval Postgraduate School Monterey, California 93943-5002	2
3.	Daniel R. Dolk, Code AS/DK Naval Postgraduate School Monterey, California 93940	2
4.	Gary K. Poock, Code OR/PK Naval Postgraduate School Monterey, California 93940	4
5.	CPT Nancy C. Browne 10370 Bridgetown Place Burke, Virginia 22015	2
6.	Norman Douglas National Communications System 8th and South Courthouse Road Arlington, Virginia 22204	2
7.	Jay Roland R. A. Rolands & Associates Corporation 500 Sloat Avenue Monterey, California 93940	1

Thesis
B8237 Browne
c.1 Speech recognition and the Telecommunications Emergency Decision Support System.